



## Analisis Regresi Logistik Biner Untuk Mengklasifikasi Penderita Hipertensi Berdasarkan Kebiasaan Merokok Di RSUD Mokopido Toli-Toli

## Binary Logistic Regression Analysis For Classification Of Hypertension Patients Based On Smoking Habits In Mokopido Toli-Toli Hospital

Misna<sup>1\*</sup>, Rais<sup>2</sup>, Iut Tri Utami<sup>3</sup>

<sup>1</sup>Jurusan Matematika Program Studi Statistika, Fakultas MIPA, Universitas Tadulako

### ABSTRACT

Hypertension is a disease caused by an abnormal increase in blood pressure, both systolic blood pressure and diastolic blood pressure. One risk factor for hypertension is smoking. This study aims to obtain the classification of hypertensive patients in Mokopido Toli-toli General Hospital using binary logistic regression. The response variable used in the study is hypertension (Y) which is categorized into two categories: hypertension and not hypertension. Predictor variables studied were smoking habits in the form of smoking time ( $X_1$ ), number of cigarettes smoked ( $X_2$ ), type of cigarettes smoked ( $X_3$ ), how to smoke cigarettes ( $X_4$ ), sex ( $X_5$ ) and age ( $X_6$ ). From the results of the study showed that the factors that influence hypertension sufferers are smoking duration ( $X_1$ ), type of cigarettes smoked ( $X_3$ ) and how to smoke cigarettes ( $X_4$ ). The binary logistic regression equation with the logit function generated is  $g(x) = -0,981 + 2,478X_1 - 2,364X_3 + 1,549X_4$ . The accuracy of classification of hypertension patients using binary logistic regression analysis is equal to 77.6%.

**Keywords:** *Binary Logistic Regression, Classification, Hypertension, Risk Habits*

### ABSTRAK

Hipertensi adalah suatu penyakit yang disebabkan oleh peningkatan abnormal tekanan darah, baik tekanan darah sistolik maupun tekanan darah diastolik. Salah satu faktor resiko terjadinya hipertensi adalah kebiasaan merokok. Penelitian ini bertujuan untuk mendapatkan klasifikasi penderita hipertensi di RSUD Mokopido Toli-toli dengan menggunakan regresi logistik biner. Variabel respon yang digunakan dalam penelitian yaitu hipertensi (Y) yang dikategorikan jadi dua kategori yaitu hipertensi dan tidak hipertensi. Variabel prediktor yang diteliti adalah kebiasaan merokok yang berupa lama merokok ( $X_1$ ), jumlah rokok yang dihisap ( $X_2$ ), jenis rokok yang dihisap ( $X_3$ ), cara menghisap rokok ( $X_4$ ), jenis kelamin ( $X_5$ ) dan usia ( $X_6$ ). Dari hasil penelitian menunjukkan bahwa faktor-faktor yang mempengaruhi penderita hipertensi adalah lama merokok ( $X_1$ ), jenis rokok yang dihisap ( $X_3$ ) dan cara menghisap rokok ( $X_4$ ). Persamaan regresi logistik biner dengan fungsi logit yang dihasilkan adalah  $g(x) = -0,981 + 2,478X_1 - 2,364X_3 + 1,549X_4$ . Nilai ketepatan klasifikasi penderita hipertensi dengan menggunakan analisis regresi logistik biner yaitu sebesar 77,6%.

**Kata Kunci :** *Regresi Logistik Biner, Klasifikasi, Hipertensi, Kebiasaan Merokok*

## LATAR BELAKANG

Hipertensi merupakan salah satu dari penyakit tidak menular yang sampai saat ini menjadi masalah kesehatan masyarakat daerah. Penyakit ini merupakan salah satu penyakit yang berbahaya karena akan membebani kerja jantung sehingga menyebabkan arteriosclerosis (pengerasan pada dinding arteri). Peningkatan tekanan darah dalam waktu lama dan tidak di deteksi sejak dini dapat menyebabkan penyakit kronik seperti retinopati, kerusakan pada ginjal, penebalan dinding jantung dan penyakit yang berkaitan dengan jantung, stroke serta kematian (Pramana, 2016)

Berdasarkan data dari Dinas Kesehatan Kabupaten Toli-toli. Angka kejadian hipertensi dari tahun 2009 sampai pertengahan tahun 2014 terus mengalami peningkatan. Dari tahun 2009 terdapat 99 kasus, pada tahun 2010 berjumlah 147 kasus, pada tahun 2011 menurun menjadi 113 kasus, pada tahun 2012 berjumlah 111 kasus, dan pada tahun 2014 kembali meningkat menjadi 616 kasus (Dinkes Toli-toli, 2014).

Banyak faktor yang menyebabkan hipertensi baik faktor yang dapat dikontrol maupun yang tidak dapat dikontrol. Faktor-faktor yang tidak dapat dikontrol antara lain usia, jenis kelamin, dan genetika (keturunan), sedangkan faktor yang dapat

dikontrol adalah merokok. Merokok merupakan salah satu kebiasaan yang dapat mempengaruhi tekanan darah. Dengan menghisap sebatang rokok akan mempunyai pengaruh besar terhadap kenaikan tekanan darah atau hipertensi sebesar 10-25 mmHg dan menambah detak jantung 5-20 kali permenit (Eirmawati dkk, 2014).

Metode statistika yang dapat digunakan untuk mengetahui faktor-faktor yang mempengaruhi penderita hipertensi di RSUD Mokopido Toli-toli adalah analisis regresi logistik biner. Analisis tersebut digunakan untuk menggambarkan hubungan antara variabel respon dengan sekumpulan variabel prediktor, dimana variabel respon bersifat biner atau dikotomis yang hanya mempunyai dua kemungkinan nilai misalnya sukses dan gagal. (Hosmer dan Lemeshow, 2000)

Pada penelitian yang dilakukan oleh Rumaenda dkk (2016) mengenai perbandingan klasifikasi penyakit hipertensi menggunakan metode regresi logistik biner dan Algoritma C4.5. Berdasarkan hasil analisis regresi logistik biner diperoleh ketepatan klasifikasi sebesar 72,5352% sedangkan pada algoritma C4.5 diperoleh ketepatan klasifikasi sebesar 64,0845%. Uji beda proporsi dihasilkan bahwa ada perbedaan signifikan antara kedua metode, sehingga dapat menyimpulkan bahwa regresi logistik

biner lebih baik dibandingkan algoritma C4.5 dalam mengklasifikasikan jenis penyakit hipertensi. Hal inilah yang mendasari penulis untuk mengkaji analisis regresi logistik biner dengan kasus yang berbeda, yaitu pengklasifikasian penderita hipertensi berdasarkan kebiasaan merokok di RSUD Mokopido Toli-toli.

## BAHAN DAN METODE

Sumber data dalam penelitian ini menggunakan data sekunder, yang diperoleh dari RSUD Mokopido Toli-toli hasil penelitian dari Muslidin pada tahun 2013 dan sampel yang digunakan sebanyak 98 pasien. Variabel respon (Y) dari penelitian ini, yaitu 1 = hipertensi dan 0 = tidak hipertensi. Variabel prediktor (X) dari penelitian ini, yaitu kebiasaan merokok berupa lama merokok (X1), jumlah rokok yang dihisap (X2), jenis rokok yang dihisap (X3), cara menghisap rokok (X4), jenis kelamin (X5) dan usia (X6).

Analisis data pada penelitian ini dilakukan dengan metode regresi logistik biner akan diolah menggunakan software SPSS versi 21.0 Berikut adalah langkah-langkah yang dilakukan untuk menganalisis data:

1. Pengumpulan data
2. Kemudian data di input kedalam SPSS 21.0
3. Menentukan model persamaan logistik biner

4. Melakukan signifikansi parameter dengan uji serentak, untuk mengetahui signifikansi parameter suatu variabel prediktor terhadap variabel respon, jika memenuhi maka dilanjutkan ke uji persial. Jika tidak memenuhi maka kembali melakukan pengumpulan data.
5. Melakukan interpretasi *odds ratio*
6. Uji kesesuaian model
7. Melakukan ketepatan klasifikasi dengan metode regresi logistik biner
8. Membuat kesimpulan dari hasil analisis.
9. Selesai

## HASIL DAN PEMBAHASAN

Berdasarkan data penelitian penderita hipertensi akan dilakukan analisis dengan menggunakan regresi logistik biner.

### 1. Regresi Logistik Biner

Setelah melakukan pengolahan data menggunakan software SPSS versi 21.0 maka diperoleh nilai estimasi untuk parameter regresi logistik biner dengan 3 variabel prediktor yang berpengaruh terhadap variabel respon adalah sebagai berikut:

Variabel	$\beta$
X <sub>1</sub>	2.478
X <sub>3</sub>	-2.364
X <sub>4</sub>	1.549
Konstanta	-0.981

Tabel 1. Estimasi Parameter Regresi Logistik Biner

Berdasarkan tabel diatas diperoleh model persamaan regresi logistic biner sebagai berikut:  $g(x) = -0,981 + 2,478X_1 - 2,364X_3 + 1,549X_4$

Pada persamaan di atas, terlihat bahwa nilai estimasi untuk lama merokok ( $X_1$ ), dan cara menghisap rokok ( $X_4$ ) bernilai positif. Ini menunjukkan bahwa hubungan variabel tersebut dengan penderita hipertensi berbanding lurus. Jika setiap nilai lama merokok dan cara menghisap rokok naik maka akan meningkatkan penderita hipertensi. Sedangkan untuk nilai estimasi jenis rokok yang dihisap ( $X_3$ ) bernilai negatif. Hal ini menunjukkan bahwa hubungan antara variabel tersebut dengan penderita hipertensi berbanding terbalik. Jika setiap nilai jenis rokok yang dihisap naik maka pada penderita hipertensi akan menurun

## 2. Uji Serentak (Uji Likelihood Ratio)

Uji serentak dilakukan untuk menguji apakah variabel prediktor secara bersama-sama berpengaruh terhadap variabel respon. Uji yang digunakan untuk menguji signifikansi model secara serentak menggunakan uji *Likelihood Ratio* yang diperoleh dengan cara membandingkan fungsi *Log Likelihood* menggunakan seluruh variabel prediktor dengan fungsi *Log Likelihood* tanpa variabel prediktor. Nilai *log-likelihood* dengan variabel prediktor dan tanpa variabel prediktor

diperoleh melalui iterasi hasil olahan menggunakan software SPSS versi 21.0. Nilai yang diperoleh masing-masing adalah sebagai berikut:

Nilai Log-likelihood dengan variabel prediktor	Nilai Log-likelihood tanpa variabel prediktor
134,835	99,696

**Tabel 2. Nilai Log-Likelihood Regresi Logistik Biner**

Berdasarkan Hosmer dan Lemeshow (2002), statistik uji yang digunakan untuk uji rasio *likelihood* adalah sebagai berikut:

$$G = -2 \ln \left[ \frac{l_0}{l_1} \right]$$

Pengujian ini menggunakan tingkat signifikan  $\alpha = 0,05$  dengan aturan keputusan  $H_0$  ditolak pada tingkat signifikan  $\alpha$  jika  $G > \chi^2_{(\alpha, v)} = G > \chi^2_{(0,05,3)}$ . Nilai statistic G berdasarkan model tersebut adalah

$$\begin{aligned} G &= -2 \ln \left[ \frac{l_0}{l_1} \right] \\ &= (-2 \ln l_0) - (-2 \ln l_1) \\ &= 134,835 - 99,696 \\ &= 35,139 \end{aligned}$$

Berdasarkan tabel *Chi-Square* dari  $\chi^2_{(0,05,3)}$  adalah 7,81. Karena nilai  $G(35,139) > \chi^2_{(0,05,3)} = 7,81$ . Maka  $H_0$  di tolak dengan tingkat kepercayaan 95%. Penolakan  $H_0$  pada tingkat signifikansi  $\alpha = 0,05$  berarti bahwa paling tidak ada satu variabel prediktor yang memiliki kontribusi yang

signifikan terhadap variabel respon. Dengan kata lain bahwa uji *likelihood ratio* menyatakan secara serentak model tersebut signifikan.

### 3. Wald (Uji Secara Parsial)

Setelah melakukan uji serentak maka langkah selanjutnya akan dilakukan pengujian signifikansi untuk masing-masing parameter dalam model diperoleh dengan cara mengkuadratkan hasil bagi estimasi parameter  $\beta_j$  dengan standar *error* estimasi parameternya. Pengujian ini menggunakan tingkat signifikan  $\alpha = 0,05$  dengan aturan keputusan  $H_0$  ditolak pada tingkat signifikan  $\alpha$  jika  $W > \chi^2_{(\alpha, db)}$  atau nilai signifikansinya lebih kecil dari  $\alpha$ . Berdasarkan hasil *output software* SPSS versi 21.0 diperoleh nilai wald untuk masing-masing variabel yang signifikan dapat disajikan pada tabel berikut:

Variabel	$\beta$	S.E	wald	sig
X <sub>1</sub>	2,478	0,606	16,721	0,000
X <sub>3</sub>	-2,364	0,615	14,799	0,000
X <sub>4</sub>	1,549	0,536	8,337	0,004
Konstanta	-0,981	0,535	3,359	0,067

Tabel 3. Hasil Uji Wald

Berdasarkan tabel 3 Pada Tabel 4.3 menjelaskan bahwa parameter yang signifikan merupakan koefisien dari variabel (X<sub>1</sub>), (X<sub>3</sub>) dan (X<sub>4</sub>), dikarenakan variabel-variabel yang mempunyai nilai sig

< 0,05. Sedangkan koefisien dari variabel X<sub>2</sub>, X<sub>5</sub> dan X<sub>6</sub> merupakan parameter yang tidak signifikan karena nilai sig > 0,05. Karena variabel X<sub>1</sub>, X<sub>3</sub> dan X<sub>4</sub> mempunyai nilai sig < 0,05 maka diputuskan  $H_0$  ditolak. Sehingga dapat disimpulkan bahwa pada tingkat signifikansi 5% penderita hipertensi yang menghisap rokok dengan waktu yang lama, jenis rokok yang dihisap dan cara menghisap rokok berpengaruh signifikan terhadap penderita hipertensi.

Setelah dilakukan uji signifikansi terhadap model diperoleh hasil bahwa variabel yang berpengaruh signifikan terhadap penderita hipertensi yaitu variabel X<sub>1</sub>, X<sub>3</sub> dan X<sub>4</sub>, sehingga model regresi logistik biner yang terbentuk adalah sebagai berikut:

$$g(x) = -0,981 + 2,478X_1 - 2,364X_3 + 1,549X_4$$

### 4. Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui kesesuaian model regresi logistik dengan membandingkan hasil pengamatan dengan nilai dugaan. Uji yang digunakan adalah uji Hosmer and Lemeshow, kriteria pengujiannya adalah tolak  $H_0$  bila nilai  $\chi^2_{hitung} < \chi^2_{tabel(db;\alpha)}$  atau nilai signifikansinya kurang dari  $\alpha = 0,05$ . Dengan bantuan *software* SPSS Versi 21.0 diperoleh nilai *chi-square* sebesar 3.003

dan signya sebesar 0,700 seperti pada tabel berikut:

$X^2_{hitung}$	Sig	$X^2_{tabel}$	df
3.003	0,700	11,07	5

**Tabel 4. Uji Kesesuaian Model**

Berdasarkan Tabel di atas menunjukkan bahwa nilai *chi-square* hitung yang didapat adalah 3,003 dengan nilai signifikansi sebesar 0,700. Karena nilai *chi square* hitung lebih kecil dari nilai *chi square* table ( $3,003 < 11,07$ ) dan nilai signifikansi lebih besar dari nilai alpha ( $0,700 > 0,05$ ), maka dapat disimpulkan bahwa model sesuai atau tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model.

#### 6. Interpretasi Koefisien (*Odds Ratio*)

Interpretasi koefisien parameter ini dilakukan untuk menggambarkan besarnya kecenderungan hasil tertentu muncul diantara kelompok yang memiliki karakter tertentu dibandingkan dengan kelompok pembanding. Setelah melakukan olahan data menggunakan *software* SPSS didapat nilai *odds ratio* masing-masing variabel prediktor yang berpengaruh terhadap variabel respon, berikut ini adalah tabel yang menunjukkan nilai *odds ratio* sebagai berikut:

Variabel	<i>Odds Ratio</i>
$X_1$	11,912
$X_3$	0,094
$X_4$	4,706

**Tabel 5. Nilai *Odds Ratio***

Interpretasi *odds ratio* masing-masing variabel adalah sebagai berikut:

##### 1. Lama merokok ( $X_1$ )

Variabel lama merokok diperoleh nilai *odds ratio* sebesar 11,912. Dari nilai tersebut dapat diartikan bahwa pasien yang menghisap rokok dengan waktu yang lama memiliki kecenderungan terkena hipertensi sebesar 11,912 kali dibandingkan dengan pasien yang menghisap rokok dengan waktu yang tidak lama yaitu kurang dari 10 tahun.

##### 2. Jenis rokok yang dihisap ( $X_3$ ).

Variabel jenis rokok yang dihisap diperoleh nilai *odds ratio* sebesar 0,094. Dari nilai tersebut dapat diartikan bahwa pasien yang merokok dengan jenis rokok filter dapat meningkatkan resiko hipertensi 0,094 kali besar dibandingkan dengan pasien yang merokok dengan jenis rokok non filter.

##### 3. Cara menghisap rokok ( $X_4$ )

Variabel cara menghisap rokok diperoleh nilai *odds ratio* sebesar 4,706. Dari nilai tersebut dapat diartikan bahwa pasien yang merokok dengan cara dalam mempunyai resiko

hipertensi 4,706 kali besar dibandingkan dengan pasien yang merokok dengan cara dangkal.

## 7. Ketepatan Klasifikasi

Persentase ketepatan klasifikasi adalah ratio antara jumlah observasi yang diklasifikasikan secara tepat oleh model dengan jumlah seluruh observasi. Dengan bantuan *software* SPSS diperoleh hasil ketepatan klasifikasi sebagai berikut:

Observasi		Prediksi	
		Tidak Hipertensi	Hipertensi
Hipertensi	Tidak Hipertensi	42	12
	Hipertensi	10	34

Tabel 6. Hasil Ketepatan Klasifikasi

Persentase ketepatan klasifikasi

$$APER = \frac{42 + 34}{98} \times 100\% = 77,6 \%$$

Perhitungan ketepatan klasifikasi dengan menggunakan persamaan (2.12) dengan metode regresi logistic biner diperoleh nilai persentase ketepatan klasifikasi berdasarkan analisis regresi logistik biner sebesar 77,6%.

Setelah melakukan pengujian statistik uji wald, diperoleh persamaan Model regresi logistik biner sebagai berikut:  $g(x) = -0,981 + 2,478X_1 - 2,364X_3 + 1,549X_4$ .

Berdasarkan model yang diperoleh terlihat bahwa nilai estimasi untuk lama merokok ( $X_1$ ), dan cara menghisap rokok ( $X_4$ ) bernilai positif. Ini menunjukkan bahwa

hubungan variabel tersebut dengan penderita hipertensi berbanding lurus. Jika setiap nilai lama merokok dan cara menghisap rokok naik maka akan meningkatkan penderita hipertensi. Sedangkan untuk nilai estimasi jenis rokok yang dihisap ( $X_3$ ) bernilai negatif. Hal ini menunjukkan bahwa jika setiap nilai jenis rokok yang dihisap naik maka pada penderita hipertensi akan menurun. Adapun model yang terbentuk dari analisis telah sesuai dan nilai persentase ketepatan klasifikasi sebesar 77,6%

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya dan rasa hormat penulis sampaikan kepada pihak Dinas Kesehatan Toli-toli yang telah memperkenankan penulis untuk mengambil data di lapangan.

## DAFTAR PUSTAKA / REFERENCES

- Dinkes Tolitoli. (2014). Profil Kesehatan "10 penyakit Terbesar". Tolitoli.
- Eirmawati, C., dkk. (2014). Hubungan antara Kebiasaan Merokok dengan Kejadian Hipertensi di RSD dr. Soebandi Jember (*Correlation Between Smoking and the Incidence of Hypertension in Department of Cardiovascular Disease RSD dr. Soebandi Jember*), Jurnal Pustaka Kesehatan, vol. 2 (no. 2), Mei 2014.
- Hosmer, D.W and Lemeshow, S. (2000). Applied Logistic Regression Second

Edition. New York: John Wiley and Sons, Inc.

Pramana, (2016). Faktor-faktor yang berhubungan dengan tingkat hipertensi di wilayah kerja puskesmas demak II Fakultas Kesehatan Masyarakat, Universitas Muhammadiyah Semarang.